Received: 1 April 2019 | Revised: 12 August 2019 | Accepted: 1 November 2019

DOI: 10.1002/aic.16866

REACTION ENGINEERING, KINETICS AND CATALYSIS

AIChE JOURNAL

Modeling of biocatalytic reactions: A workflow for model calibration, selection, and validation using Bayesian statistics

Ina Eisenkolb[1] | Antje Jensch[1] | Kerstin Eisenkolb[1] | Andrei Kramer[2] | Patrick C. F. Buchholz[3] | Jürgen Pleiss[3] | Antje Spiess[4] | Nicole E. Radde[1]

[1]Institute for Systems Theory and Automatic Control, University of Stuttgart, Stuttgart, Germany

[2]Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden

[3]Institute for Biochemistry and Technical Biochemistry, University of Stuttgart, Stuttgart, Germany

[4]Institute for Biochemical Engineering, Technical University Braunschweig, Braunschweig, Germany

Correspondence
Nicole E. Radde, Institute for Systems Theory and Automatic Control, University of Stuttgart, Pfaffenwaldring 9, D-70569 Stuttgart, Germany.
Email: nicole.radde@ist.uni-stuttgart.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: PL145/16-1, EXC 2075-390740016, EXC 310/2

Abstract

We present a workflow for kinetic modeling of biocatalytic reactions which combines methods from Bayesian learning and uncertainty quantification for model calibration, model selection, evaluation, and model reduction in a consistent statistical framework. Our workflow is particularly tailored to sparse data settings in which a considerable variability of the parameters remains after the models have been adapted to available data, a ubiquitous problem in many real-world applications. Our workflow is exemplified on an enzyme-catalyzed two-substrate reaction mechanism describing the symmetric carboligation of 3,5-dimethoxy-benzaldehyde to (*R*)-3,3′,5,5′-tetramethoxybenzoin catalyzed by benzaldehyde lyase from *Pseudomonas fluorescens*. Results indicate a substrate-dependent inactivation of enzyme, which is in accordance with other recent studies.

KEYWORDS

carboligation, enzyme kinetics, Markov chain Monte Carlo, parameter estimation, profile likelihood, residual analysis, thiamine-diphosphate-dependent enzymes

1 | INTRODUCTION

Modeling is a difficult task with many challenges. A good model is predictive and helps to get deeper insight into the described system or phenomenon by, for example, explaining underlying mechanisms or giving raise to nonobvious hypotheses which can be tested in a subsequent step. In many applications, building a good model renders it necessary to adapt the model's granularity to the available data and also to the kind of scientific questions the model should answer.

For (bio-)chemical reaction networks, as considered here on an example of an enzyme-catalyzed two-substrate conversion, standard modeling approaches based on chemical reaction kinetics exist. In addition, several workflows guide the researcher through a sequence or iteration of experiments, parameter estimation and model quality assessment as well as model refinement and/or experimental design steps.[1-5] These models are usually available in parameterized form, and parameter estimation is formulated as an optimization problem in which the model behavior is calibrated to experimental data. We often face the problem that the data sets do not contain enough information for the parameters to be uniquely identified. This is either caused by structural nonidentifiabilities, that is, the kind of data used for estimation do generally not allow for a unique parameter identification regardless of a particular realization and measurement noise, or by practical nonidentifiabilites, which can occur in a setting in which given data are sparse or very noisy. Methods based on point estimates and local approximations cannot be used in these cases. Structural identifiability can be tested by algebraic methods, which are often only applicable to small or medium size models or need at least a reference value and suitable neighborhood.[6] Regularization methods can be used to convert ill-posed into well-posed problems for particular

Ina Eisenkolb and Antje Jensch contributed equally to this study.

problem settings.[7] Suitable methodology to investigate practical identifiability comprises profile likelihood analysis[8] and Bayesian methods.[9] The later also allow for a consistent uncertainty quantification from data to model parameters to model predictions for any quantity of interest.

In addition to nonidentifiable parameters, sometimes the reaction mechanism is not known in full detail, the reaction kinetics is influenced by its environment, or the reaction system is embedded in a larger reaction network with yet uncharacterized crosstalk effects. Then, the model is not completely specified, resulting in problems such as structure identification or model selection, which are usually even more difficult than parameter estimation for a single model.[10]

Here, we introduce a modeling workflow for parameter estimation, model selection, model reduction, and validation based on Bayesian statistics, which is particularly tailored for consistent uncertainty quantification, and compare it to a similar workflow which uses local methods.[11] Moreover, we discuss different ways to visualize outcomes of individual steps in the workflow. We regard such a workflow as a prerequisite for an automated data and model management system, which makes modeling results transparent and reproducible and facilitates standardization of processes.

Our approach is exemplified on an enzyme-catalyzed two-substrate reaction mechanism describing the symmetric carboligation of 3,5-dimethoxy-benzaldehyde (3,5-DMBA) to (*R*)-3,3′,5,5′-tetramethoxybenzoin ((*R*)-3,3′,5,5′-TMB) catalyzed by benzaldehyde lyase (BAL) from *Pseudomonas fluorescens* (PDB accession 2AG0), for which we compare two competing model hypotheses. Results confirm previous recent findings about this reaction mechanism[4,5,11] and illustrate that global methods such as sampling-based analysis provide superior insights into underlying parameter dependencies compared to local approximations.

2 | RESULTS

2.1 | Modeling workflow

Our proposed modeling workflow is depicted in Figure 1. We start with an experimental data set and a set of initial models, here exemplified with models 1 and 2. These models are calibrated independently by using sampling-based Bayesian approaches. For model selection, we compare model fits and introduce additional methods, which make use of residual and parameter identifiability analysis, to judge overall plausibility of the models. If both models give satisfactory results, one can use model-based experiment design to suggest further experiments which help to discriminate between both models, leading to

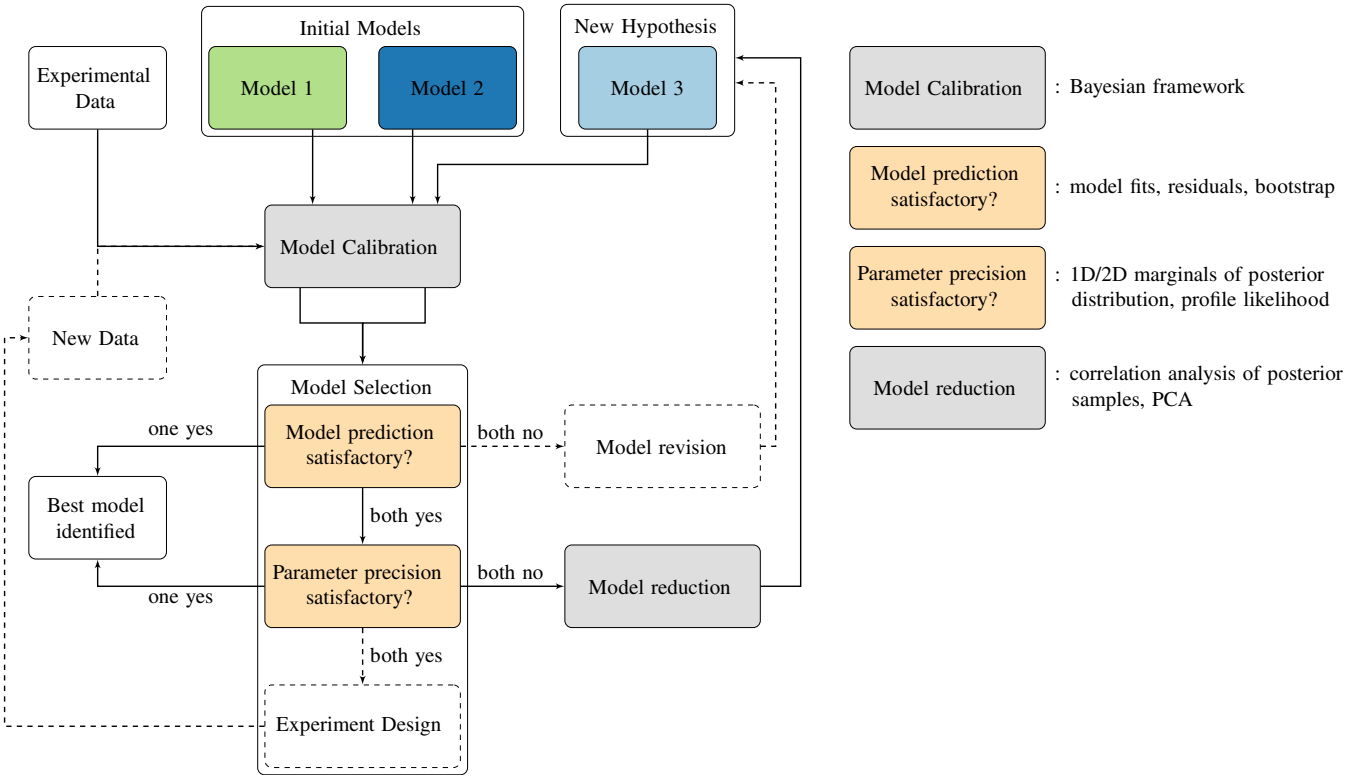


FIGURE 1 Modeling workflow. Different parameterized initial model hypotheses are calibrated independently to experimental data. For model selection, we evaluate model predictions and parameter precision. If more than one model hypothesis gives satisfactory results, model-based experiment design is used to discriminate between these alternatives. If all models fail, the models are revised either by including expert knowledge or by model reduction techniques, leading to new model hypotheses, which enter the workflow from scratch. Color code of the model blocks is used for subsequent presentation of results. Dashed parts indicate workflow blocks not applied in this study. Methods employed for each block are listed on the right [Color figure can be viewed at wileyonlinelibrary.com]

new experimental data. Methods for this are available and are, for example, based on Fisher information matrices[12] or on optimizations of output variances.[13] In case that none of the two models provide satisfactory fits, the model has to be revised, leading to a new model hypothesis. If the fits are good but parameter precision is not satisfactory, we employ model reduction techniques based on the sampled parameters, which also leads to a new model hypothesis that can be compared in the same way. The individual workflow steps are explained in more detail for the particular example in the following. Color codes for the three models in Figure 1 are reused in the simulations of the respective models in this study. We do not explicitly consider model-driven experiment design methods and model revision here, which is indicated by dashed lines and boxes in Figure 1, since standard methods are available for this and model revision is usually done manually, for example, via including expert knowledge and/or more details about the process at hand.

2.2 | Experimental time course data and competing modeling approaches

We use experimental data obtained from Zavrel et al.,[11] which describe the symmetric carboligation of 3,5-DMBA resulting in (*R*)-3,3′,5,5′-TMB catalyzed by BAL (Figure 2a). The conversion of substrate was measured as time courses in nine experiments, which differ in initial substrate and/or enzyme concentrations (Figure 2b). The range in which the initial substrate concentration A_0 could be varied was bounded by limited product solubility. Initial enzyme concentration E was either set to 4.17×10^{-5} mM or 8.33×10^{-5} mM, and the initial product concentration P_0 was 0 mM in all experiments. Equidistant measurement points with $\Delta t = 6$ s were taken until equilibrium was reached, 2,777 data points in total were available.

Two initial model hypotheses (models 1 and 2) based on the mechanistic kinetic models derived in Zavrel et al.[11] are calibrated and compared in a first step. The enzyme reaction of BAL follows an ordered bi-uni reaction mechanism since two (identical) substrate molecules are converted to one product molecule. Both models are based on the same reaction mechanism, which is shown with its microreaction steps in Figure 2c.[14] Here, we have used *A* and *B* for the two identical substrates, which bind with different affinities to the enzyme (donor-acceptor principle).[15]

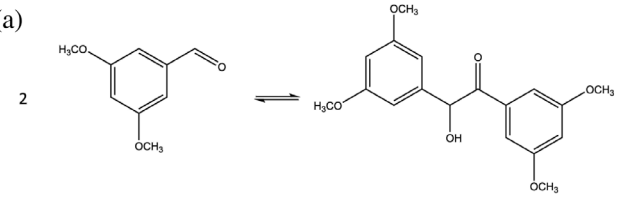
The macrokinetic model reads[11]

$$v = -\frac{1}{2}\frac{dA}{dt} \tag{1a}$$

$$= E \cdot \frac{\frac{k_{catf}}{K_{iA}K_{mB}}\left(A^2 - \frac{P}{K_{eq}}\right)}{1 + \frac{A}{K_{iA}}\left(1 + \frac{K_{mA}}{K_{mB}}\right) + \frac{A^2}{K_{iA}K_{mB}} + \frac{P}{K_{mP}} + \frac{AP}{K_{mP}K_{iB}}}, \tag{1b}$$

with seven parameters that are listed in Figure 2d. Due to the fact that only six microreaction constants exist (Figure 2c), we employ the functional relation.[11]

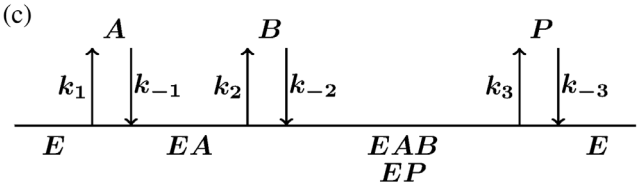
(a)



(b)

Experiment	A_0 [mM]	P_0 [mM]	E_0 [mM]	Duration [s]
1	3.00	0	$8.33 \cdot 10^{-5}$	1495
2	2.75	0	$8.33 \cdot 10^{-5}$	1495
3	2.50	0	$8.33 \cdot 10^{-5}$	1495
4	2.25	0	$8.33 \cdot 10^{-5}$	1453
5	2.00	0	$8.33 \cdot 10^{-5}$	697
6	1.50	0	$8.33 \cdot 10^{-5}$	697
7	3.00	0	$4.17 \cdot 10^{-5}$	2995
8	2.75	0	$4.17 \cdot 10^{-5}$	2995
9	2.50	0	$4.17 \cdot 10^{-5}$	2995

(c)



Model 1: $\boldsymbol{E} = \boldsymbol{E_0}$ Model 2: $\frac{d}{dt}\boldsymbol{E} = -\boldsymbol{k_{ins}AE}$

(d)

Param	Model	Meaning	Unit
k_{catf}	1,2	Maximum turnover number	s^{-1}
K_{eq}	1,2	Equilibrium constant	mM^{-1}
K_{mA}	1,2	Michaelis constant for A (donor) to E	mM
K_{mB}	1,2	Michaelis constant for B (acceptor) to EA	mM
K_{mP}	1,2	Michaelis constant for P to E	mM
K_{iA}	1,2	Dissociation constant of EA	mM
K_{iB}	1,2	Dissociation constant of EAB	mM
k_{inS}	2	Inactivation constant of E	$mM^{-1}s^{-1}$

FIGURE 2 Bi-uni reaction mechanism specifics according to Reference 14. (a) Reaction scheme for the symmetric carboligation of 3,5-dimethoxy-benzaldehyde (3,5-DMBA) to (*R*)-3,3′,5,5′-tetramethoxybenzoin ((*R*)-3,3′,5,5′-TMB) catalyzed by benzaldehyde lyase (BAL). (b) Initial experimental conditions and time periods until reaching equilibrium. (c) First, the enzyme-substrate complex *EA* forms by the binding of substrate *A* (donor) to enzyme *E*. This is followed by the binding of *B* (acceptor) resulting in the ternary *EAB* complex. Finally, the enzyme *E* releases the product *P*. (d) Parameters of Models 1 and 2. The parameter k_{inS} only applies to Model 2

$$K_{iB} = \frac{K_{mB}K_{iA}}{K_{mA}\left[1 - \left(\frac{K_{mA}}{K_{iA}} - 1\right)\frac{K_{mP}}{K_{eq}K_{mB}K_{iA}}\right]}. \tag{2}$$

We refer to this model as model 1, with parameter vector $\theta^1 = (k_{catf}, K_{eq}, K_{mA}, K_{mB}, K_{mP}, K_{iA})$. Model 2 includes in addition a substrate-dependent enzyme inactivation, which is described by mass action kinetics,

$$\frac{dE}{dt} = -k_{inS} \cdot A \cdot E, \tag{3}$$

such that Model 2 comprises one additional parameter k_{inS} and hence $\theta^2 = (k_{catf}, K_{eq}, K_{mA}, K_{mB}, K_{mP}, K_{iA}, k_{inS})$. The theory that the enzyme

BAL is inactivated by the substrate 3,5-DMBA is in accordance with previous studies, in which an inactivation of a decarboxylase by benzaldehydes has also been reported.[16-18]

2.3 | Both models capture experimental data but differ in parameter values and confidence bounds

For model calibration, we use a Bayesian approach, which provides a consistent description of all quantities of interest in terms of probability distributions and thus allows to transform variability in measurements into uncertainties in parameters and, ultimately, model predictions. We do this separately for all initial models. Such statistical approaches require a stochastic model in order to define a likelihood function. Therefore, we embed the differential equation system into a stochastic error model. The resulting model generally captures besides the dynamics of the system also the noise characteristics, which can be integrated into the objective function, and allows to analyze model fits and uncertainty in terms of comparisons of probability distributions, for which measures from statistics and information theory are available. Several variants of such an error model exist in the literature, and additive normal errors, multiplicative log-normal errors or a mixture of both are among the most frequently used.[19] A proper choice of parameters for these error models is another task. In some cases, they can be fixed beforehand, for example, by data preprocessing (see for example, Jensch et al.[20]), in other cases they have to be estimated together with the other model parameters. Often it makes sense to use a shrinkage approach or to partly pool parameters such as SD for the same model output within the same experiment. As pooling reduces the number of parameters, it is often used for computational reasons.

Here, we employ a multiplicative error model according to

$$y_A^j(t_k) = z_A^j(t_k,\theta)\cdot\epsilon \quad \epsilon \sim \log N\left(0,\sigma^2\right), \tag{4}$$

where $z_A^j(t_k,\theta)$ refers to the solution of the differential equation for substrate A at time point t_k, $k \in 1, \ldots, T$ in experiment $j \in 1, \ldots, 9$ and $y_A^j(t_k)$ denotes the respective noise-corrupted measurement.

The resulting likelihood function for experimental data set D reads

$$L_D(\theta) = \prod_{j=1}^{9}\prod_{k=1}^{T}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(\log z_A^j(t_k,\theta) - \log y_A^j(t_k)\right)^2}{2\sigma^2}\right). \tag{5}$$

A Bayesian approach requires a prior distribution $\pi(\theta)$ on the parameters, for which we choose a uniform distribution on the log-transformed parameters. This transformation is usually employed in cases where the order of magnitude of the unknown parameters is not known a priori since it enables to cover several orders of magnitude and obeys Benford's law according to which the mantissa of logarithms of numbers are equally distributed.[21,22] In addition, it maps positive parameters onto the entire set of real numbers, and in this way we get rid of this positivity assumption as a constraint. This log transformation has been shown to be highly advantageous for parameter estimation in systems biology models, see for example, Kreutz[23] and Villaverde et al.[24] Of note is here that probability distributions are generally not invariant for such nonlinear transformations, and results are different from choosing a uniform prior for the nontransformed parameters. An unbounded uniform measure corresponds to an improper prior distribution, resulting in the posterior distribution being proportional to the likelihood function.

For implementation, we use a proper prior distribution on the parameters by choosing reasonable bounds. This is usually done in an adaptive way. We cover a large range initially, solve the optimization problem to find the maximum a posteriori (MAP) estimator, and then adapt the boundaries such that they enclose this estimator and comprise most of the probability mass. The MAP estimator is thus defined as

$$\hat{\theta}^{\mathrm{MAP}} = \mathrm{argmin}_\theta - \log L_D(\theta) \text{ subject to } \theta \in [\theta_\mathrm{l}, \theta_\mathrm{u}] \tag{6}$$

with corresponding objective function value J_{opt}

$$J_{\mathrm{opt}}\left(\hat{\theta}^{\mathrm{MAP}}\right) = -\log L_D\left(\hat{\theta}^{\mathrm{MAP}}\right). \tag{7}$$

The posterior distribution then reads

$$p(\theta|D) = \frac{\pi(\theta)L_D(\theta)}{p(D)}, \quad p(D) = \int \pi(\theta)L_D(\theta)d\theta\,. \tag{8}$$

This distribution is investigated by sampling, which circumvents evaluating the integral $p(D)$. Direct sampling techniques that provide independent samples are usually not efficient and Markov Chain Monte Carlo (MCMC) sampling is used instead. Different sampling schemes exist for this purpose, which are discussed, for example, in Weber et al.[10] and Kramer et al.[25,26] For technical details about the sampling scheme that we use and convergence tests we refer to supporting Section 1.

Once a representative sample of the posterior distribution is available, distributions of interest, for example, marginal and posterior predictive distributions (PPDs) can be analyzed. Marginal distributions for parameters of interest can be obtained by integrating the joint probability distribution over all other parameters. In case of 1D and 2D marginal distributions, this reads

$$p(\theta_i|D) = \int_{\theta_{m\neq i}} p(\theta|D)d\theta_m \quad \text{and} \quad p\left(\theta_i,\theta_j|D\right) = \int_{\theta_{m\neq i,\,j}} p(\theta|D)d\theta_m \tag{9}$$

These PPDs or, generally, PPDs for any quantity of interest $\tilde{y}$ can be inferred via Monte Carlo integration,

$$p(\tilde{y}|D) = \int p(\tilde{y},\theta|D)d\theta \tag{10a}$$

$$= \int p(\tilde{y}|\theta,D)p(\theta|D)d\theta \tag{10b}$$

$$= \int p(\tilde{y}|\theta)p(\theta|D)d\theta \tag{10c}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} p(\tilde{y}|\theta_i)\,, \tag{10d}$$

where θ_i are N samples from the posterior distribution $p(\theta|D)$. For parameter marginals, that is, $\tilde{y}$ is a subset of the parameters, this translates into just considering the respective entries in the sampled parameter vectors and using these to estimate marginal densities. Since the θ_i have been generated from a Markov process, the series is usually autocorrelated, which has to be taken into account when estimating the variance of the measure of interest. The variable $\tilde{y}$ can, for example, also be a state $y(t^*)$ at a particular time point t^* or a waiting time until a particular event occurs.

If the measurement noise σ^2 in Equation (4) is small, interquartile ranges of this distribution can directly be approximated by using the posterior sample to simulate a bundle of trajectories and using these to empirically estimate these ranges.

Figure 3 shows model fits for three exemplary time series for Model 1 (left) and Model 2 (right). Data are represented as dots, solid lines indicate MAP estimators $\hat{\theta}^{\text{MAP}}$. Empirical interquartile ranges of trajectory bundles are omitted here since they form very narrow bands around the maximum likelihood values, as suggested by the data.

Model fits of the remaining six experiments are shown in Figure S1. Visually, there is not much difference between the two model variants in terms of fit quality.

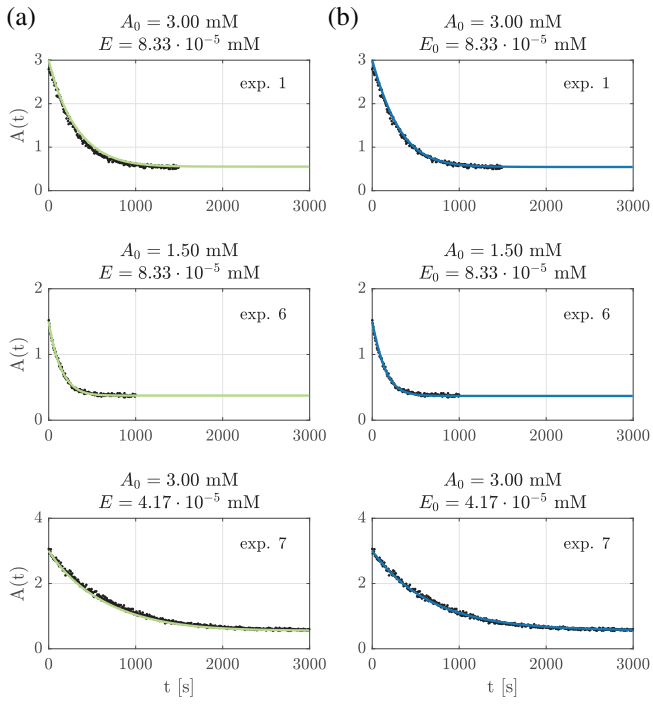


FIGURE 3 Calibrated model trajectories for (a) Model 1 and (b) Model 2. A comparison of model trajectories simulated with the maximum a posteriori (MAP) estimator $\hat{\theta}^{\text{MAP}}$ and experimental time courses for three exemplary initial conditions (Experiments 1, 6, and 7) [Color figure can be viewed at wileyonlinelibrary.com]

The MAP estimators for Models 1 and 2 are listed in Figures 4a and 5a, respectively, and are indicated as white dots in Figures 4b and 5b. Figures 4b and 5b visualize estimated 2D parameter marginals $p(\theta_i, \theta_j|D)$ (Equation (9)) of the posterior distributions for Models 1 and 2 as 2D scatterplots (lower right half) and corresponding binning plots on a hexagonal lattice (upper left half). 1D Parameter marginals (Equation (9)) are depicted below.

MAP estimates of both models are similar only for the parameters k_{catf} and K_{eq}. All other parameters differ significantly by several orders of magnitude, showing that estimated parameter values strongly depend on model equations and thus have to be handled with care. This is a phenomenon which was also observed in Buchholz et al.[5]

As indicated by the 1D marginal distributions for Model 1, the parameters k_{catf} and K_{eq} are well identifiable. The parameter K_{iA} has a broader distribution, and those of K_{mA}, K_{mB}, and K_{mP} cover a wide range and are not identifiable. Correlations with absolute values of coefficients larger than 0.5 only exist between K_{iA} and K_{mA} (ρ = 0.67) and between K_{iA} and K_{mB} (ρ = −0.69).

For Model 2, k_{catf}, K_{eq}, K_{mA}, K_{mP}, and k_{ins} are well identifiable, while K_{mB} and K_{iA} have broader distributions. Moreover, the parameters are much more correlated. Besides the two correlations which are also present in Model 1, $\rho(K_{\text{iA}}, K_{\text{mA}})$ = 0.72 and $\rho(K_{\text{iA}}, K_{\text{mB}})$ = −0.96, the parameters K_{iA} and K_{mP}, K_{mP}, and K_{mA}, K_{mP}, and K_{mB} as well as K_{mA} and K_{mB} are also correlated with a coefficient $\rho > 0.5$.

We note here that these 1D and 2D parameter marginals clearly show that a FIM-based local analysis, as presented in Reference 11, comes to its limits. Confidence bounds are estimated in this analysis by approximating the 2D densities with bivariate normal distributions centered around $\hat{\theta}^{\text{MAP}}$, and corresponding 1D marginals, which are also normally distributed. While this might still be a reasonable approximation for some parameters of Model 1, in particular the well-identifiable parameters k_{catf} and K_{eq}, it cannot describe the nonlinear relationships between parameters of Model 2.

In summary, although Model 2 has more parameters, these parameters are much better identifiable than those of Model 1, which indicates that Model 2 is better suited to describe the data. Estimation of confidence bounds via profile likelihoods analysis,[8,27-29] as depicted in Figure 6, confirms these results. Respective confidence interval lengths are listed in Figure 6c. Changes of other parameters along the profile curves are shown in Figure S2.

2.4 | Model comparison suggests substrate-dependent enzyme inactivation

Residual analysis is a way to judge whether the deviations between the progress curves produced by $\hat{\theta}^{\text{MAP}}$ and the data are in the range of noise. Therefore, we use our trained stochastically embedded differential equation model (1b) to simulate experimental data and compare residuals with those observed in real experiments. In an ideal setting, for example, in an in silico study where the calibrated model was also used to generate artificial data and all parameters are identifiable, both residual sets are stochastically indistinguishable. When

(a)

	k_{catf} [s^{-1}]	K_{eq} [mM^{-1}]	K_{mA} [mM]	K_{mB} [mM]	K_{mP} [mM]	K_{iA} [mM]
$\log_{10}$-value	1.69	0.60	−7.24	−6.03	0.72	−2.29
real value	49.01	4.01	$5.72 \cdot 10^{-8}$	$9.31 \cdot 10^{-7}$	5.23	0.01
$\log_{10}$-bounds	[1.5, 1.9]	[0.4, 0.8]	[−8, 0]	[−8, 0]	[−5, 5]	[−5, 5]

(b)

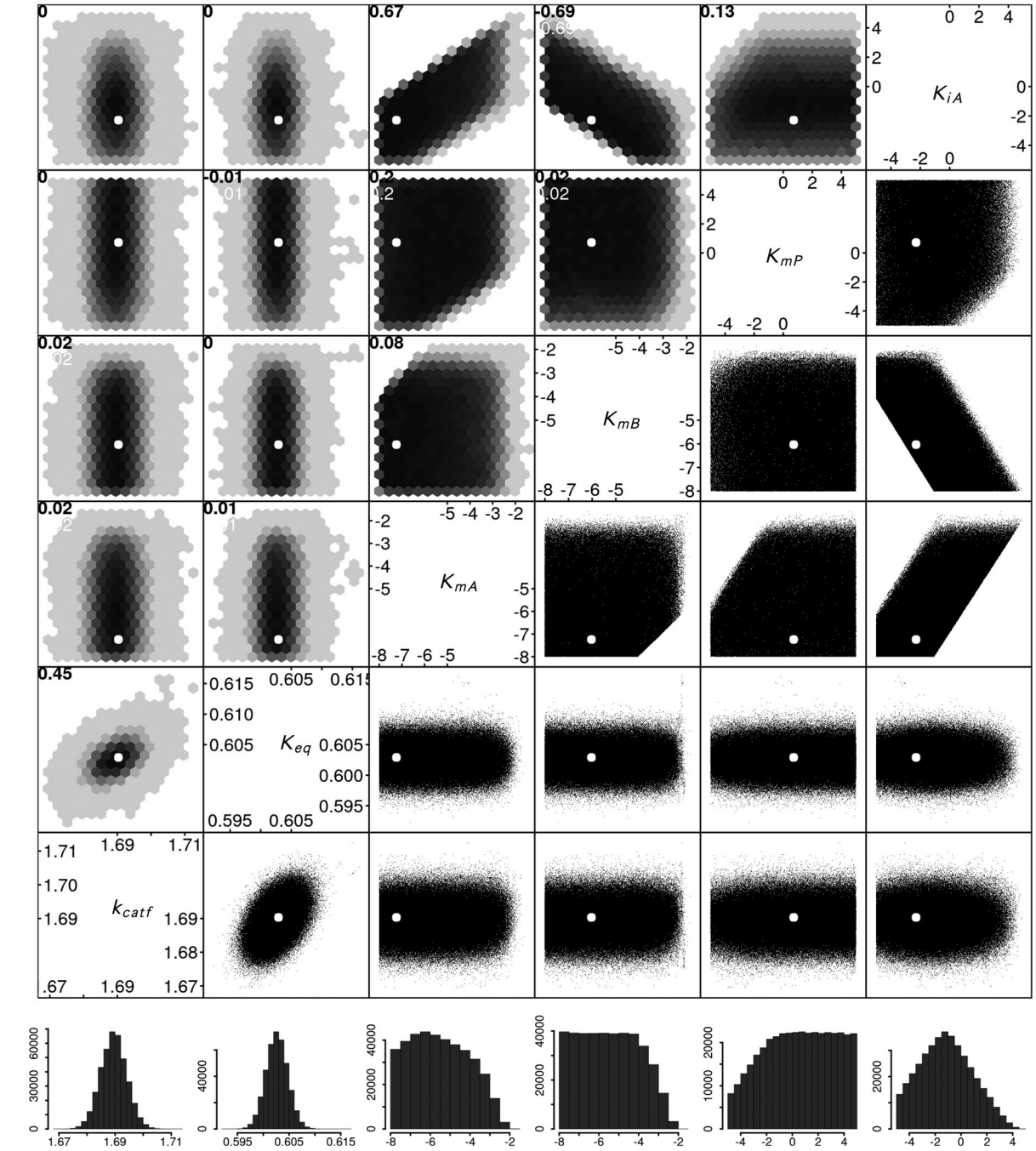


FIGURE 4 Maximum a posteriori (MAP) estimates and 2D marginal posterior densities for Model 1. (a) Best parameter vectors $\hat{\theta}^{MAP}$ as well as upper and lower parameter bounds of the Markov Chain Monte Carlo (MCMC) sampling for Model 1. (b) Parameter scatterplots (lower right half) and 2D histograms with hexagonal bins (upper left half) for the MCMC samples of Model 1. MAP estimates $\hat{\theta}^{MAP}$ are indicated by white dots. 1D marginal densities are shown below (for parameters named in the column above). Correlation coefficients ρ are shown in the upper left corner of the hexbin plots. Units are as listed in (a)

(a)

	k_{catf} [s^{-1}]	K_{eq} [mM^{-1}]	K_{mA} [mM]	K_{mB} [mM]	K_{mP} [mM]	K_{iA} [mM]	k_{insS} [$mM^{-1}s^{-1}$]
$\log_{10}$-value	1.66	0.62	−0.86	−3.56	−0.64	2.53	−3.59
real value	46.22	4.15	0.14	$2.77 \cdot 10^{-4}$	0.23	333.38	$2.60 \cdot 10^{-4}$
$\log_{10}$-bounds	[0, 3]	[−1, 2]	[−5, 0]	[−8, 0]	[−5, 3]	[−5, 5]	[−4.5, −2.5]

(b)

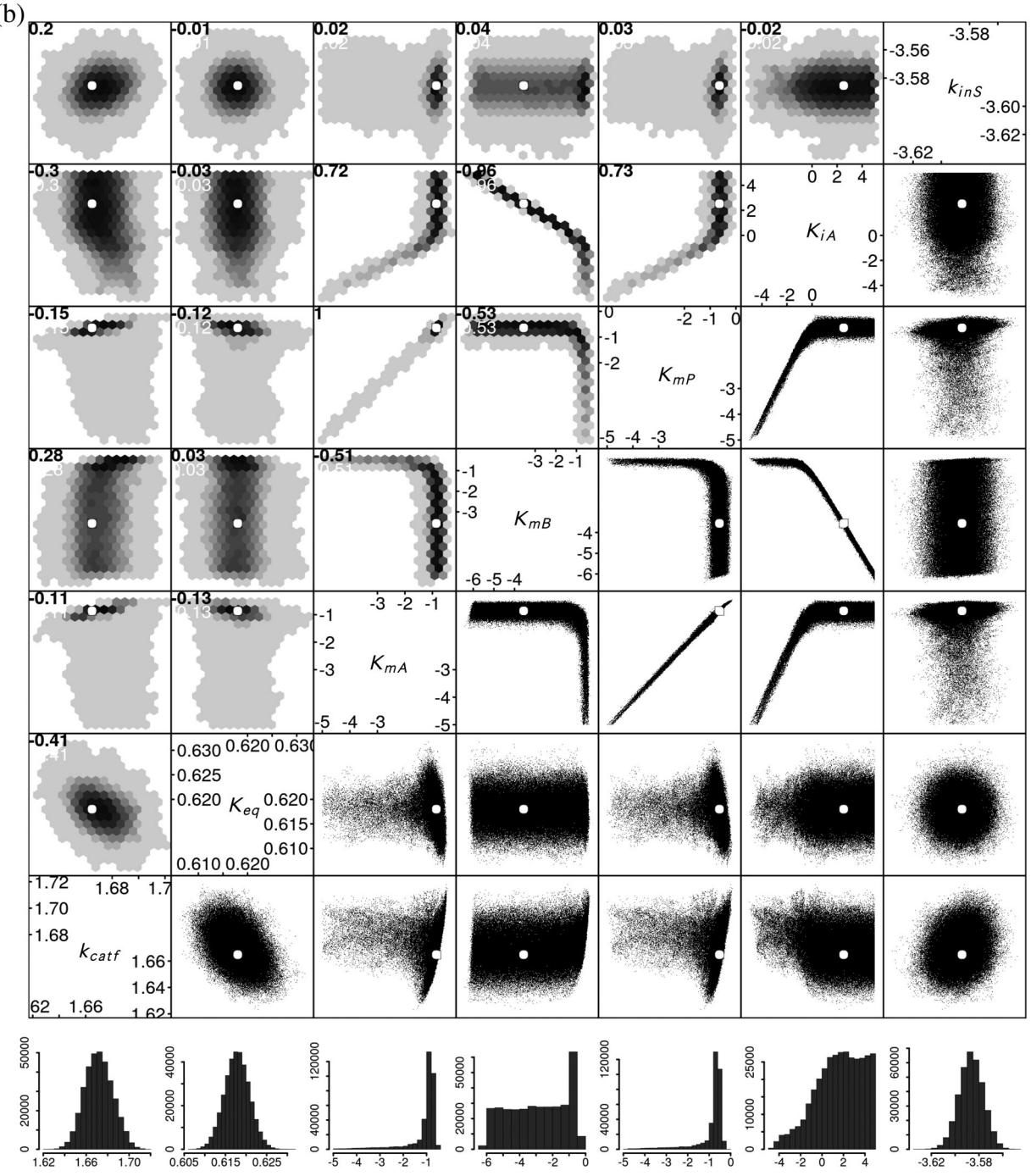


FIGURE 5 Maximum a posteriori (MAP) estimates and 2D marginal posterior densities for Model 2. (a) Best parameter vectors $\hat{\theta}^{\text{MAP}}$ as well as upper and lower parameter bounds of the Markov Chain Monte Carlo (MCMC) sampling for Model 2. (b) Parameter scatterplots (lower right half) and 2D histograms with hexagonal bins (upper left half) for the MCMC samples of Model 2. MAP estimates $\hat{\theta}^{\text{MAP}}$ are indicated by white dots. 1D marginal densities are shown below (for parameters named in the column above). Correlation coefficients ρ are shown in the upper left corner of the hexbin plots. Units are as listed in (a)

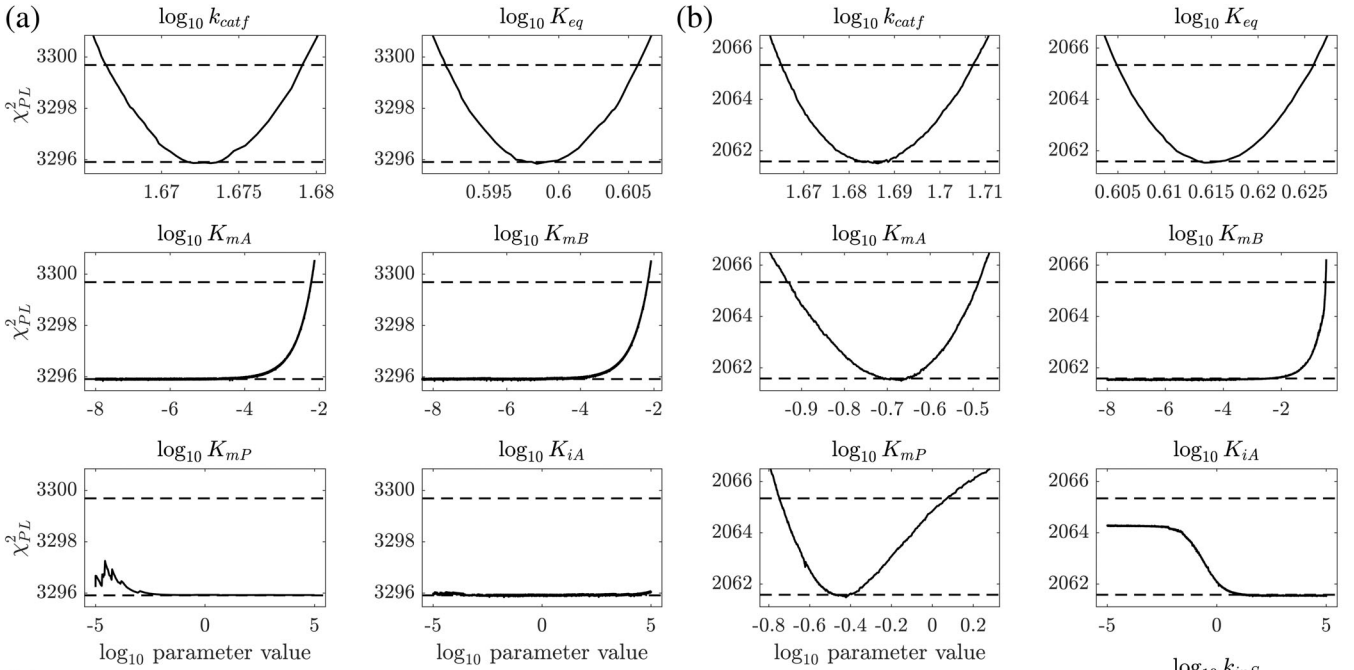
working with experimental data, this might also be the case if the stochastic noise is large compared to the deviation and residuals are dominated by this stochastic part. In case of small noise, however, the difference between the sets of residuals should be visible.

Figure 7 shows a comparison of both sets of residuals: for Models 1 (left) and 2 (right). Residuals of simulated and experimental data are depicted in red and green (Model 1) and red and blue (Model 2), respectively.

Looking at Model 1, the long-term behavior, which describes the equilibrium of the reaction, is well captured and both residual sets are visually almost indistinguishable for $t \geq 500$ s. However, this is different for some transients. In particular, residuals of the experimental data are clearly above those of the simulated data for large enzyme concentrations, and below for small enzyme concentrations, respectively. Figure S3a shows that this is consistently true for all five substrate concentrations. Interestingly, these effects average out when taking the residual means of all experiments. For Model 2, ranges of both residual sets consistently overlap in almost all experiments, indicating again that the experimental data and especially their noise characteristics are better captured by Model 2.

We employ parametric bootstrap[30] for an aggregated analysis of the model fit. We use the calibrated model to simulate responses under the same input conditions and noise as for the data sets used for parameter estimation. For each of these output sets, the likelihood function value, given the calibrated model, is evaluated. The maximum likelihood value of the experimental data is compared to the distribution of those in silico results (represented by box and whisker plots). If it lies below the fifth percentile, experimental data are far more likely than most of the simulated data, which is an indication for overfitting. Similarly, a value above the 95th percentile means that most simulated data sets are in better accordance with model predictions than the experimental data set and the model might lack relevant structural information. Results of this analysis for Models 1 (green) and 2 (blue) are depicted in the last row of Figure 7 for an average of all experiments.

The difference between the two models is small in this analysis for all experiments combined as well as for all nine experiments (Figure S4). The MAP estimator of both models lies within the major mass of the empirical distribution and thus both models cannot be rejected based on this analysis. Of note is that this analysis has to be interpreted with care in two respects. First, the resulting empirical distribution is highly sensitive to the choice of the error model and its parameters. Second, the effects of underfitting and overfitting of individual experiments often average out in the overall averaged



(c)

	k_{catf} [s^{-1}]	K_{eq} [mM^{-1}]	K_{mA} [mM]	K_{mB} [mM]	K_{mP} [mM]	K_{iA} [mM]	k_{insS} [$mM^{-1}s^{-1}$]
Model 1							
$\log_{10}$-value	0.01	0.01	inf	inf	inf	inf	–
real value	1.38	0.13	inf	inf	inf	inf	–
Model 2							
$\log_{10}$-value	0.04	0.02	0.45	inf	0.82	inf	0.05
real value	4.72	0.20	0.21	inf	0.99	inf	$3.00 \cdot 10^{-5}$

FIGURE 6 Profile likelihood analysis for Models 1 and 2. Profile likelihood analysis for parameters of (a) Models 1 and (b) 2 and (c) profile likelihood 95% confidence interval lengths of the respective parameters

bootstrap statistics. Thus, parametric bootstrap as introduced here might be a way to reject models especially in cases where the error model is well known.

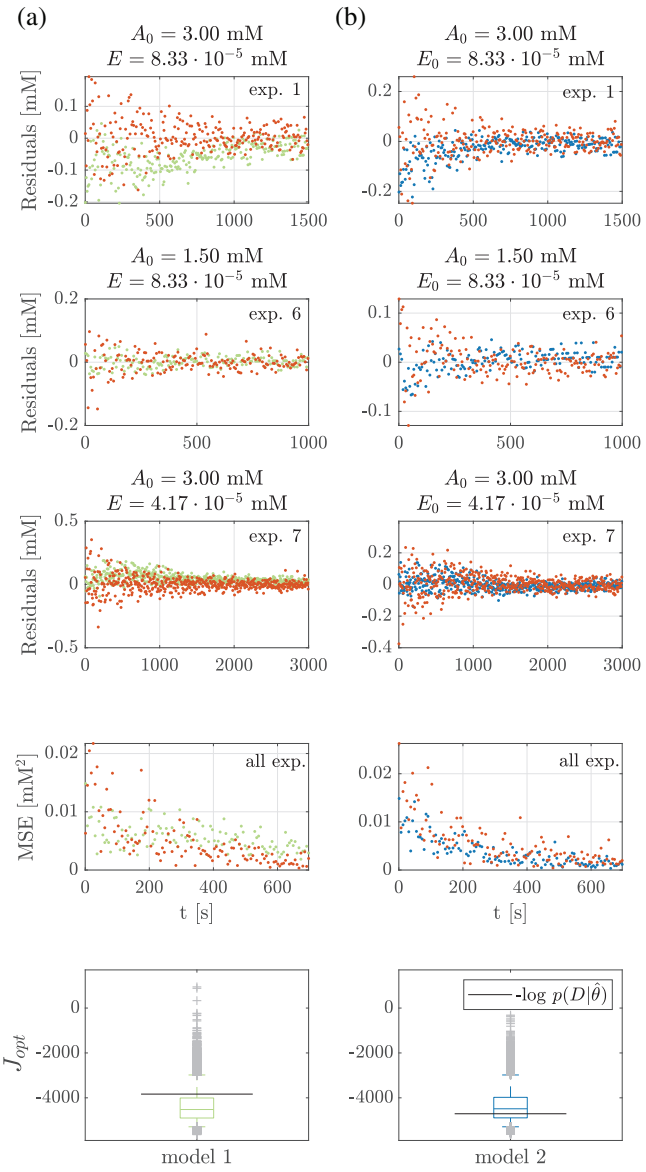


FIGURE 7 Residual and bootstrap analysis for Models 1 and 2. Red dots indicate residuals simulated with the maximum a posteriori (MAP) estimator and the proposed multiplicative noise model for three representative experiments, plotted against time. (a) Green (Model 1) and (b) blue (Model 2) dots are the residuals resulting from a comparison between the MAP trajectory and the experimental data. The respective mean squared error (MSE) is shown in the fourth row. In the last row, a distribution of J_{opt} has been generated by a bootstrap analysis, in which many data sets and respective likelihood function values were simulated with (a) Model 1 and (b) Model 2 for all experiments. Resulting data were used to estimate the median, the 25th and 75th percentiles (colored boxes) and the 5th and 95th percentiles (lower and upper adjacent values, respectively). Outliers are depicted as crosses. Horizontal lines depict J_{opt} values of the real experiments

Overall, Model 2 is favorable according to our analysis, which supports the hypothesis of a substrate-dependent enzyme inactivation.

2.5 | Model reduction based on statistical analysis

Equipped with a model which gives satisfactory predictions, but whose parameters are not completely identifiable, we decided to apply model reduction techniques to Model 2 according to our workflow (Figure 1). Model reduction is a broad field, and many different techniques are on the market. For larger chemical reaction networks, time-scale separation techniques such as quasi-steady state or rapid equilibrium approximations are sensible approaches.[31-33] Model-order reduction techniques can be applied to reduce the computational complexity in numerical simulations. Since we are interested in improving parameter identifiability, and we already observed several linear correlations between pairs of parameters in Model 2 (Figure 5b), we applied principal component analysis to the parameter samples of Model 2 in order to investigate whether subgroups of linearly related parameters exist. The following correlations were found:

$$0 = 0.7819K_{mA} - 0.7489K_{mP}\,, \tag{11}$$

$$0 = -0.3682K_{mA} + 0.3800K_{mB} + 0.3733K_{iA}\,. \tag{12}$$

The scatterplots in Figure 8 show the 2D (left) and 3D (right) correlations. In order to be able to compare to results in Zavrel et al.[11] and Buchholz et al.,[5] these linear relations were used to eliminate the parameters K_{mP} and K_{iA}.

The parameters of the reduced model (Model 3) were sampled and Figure S2b shows that all parameters except K_{inS} are well identifiable with small variances. Interestingly, model reduction led to an almost completely flat distribution of K_{inS}.

Results of the fit analysis are shown in Figure 9. Model fits look reasonable in terms of time series fits (Figure 9a). The visual analysis of the residuals in Figure 9b, however, reveals systematic deviations

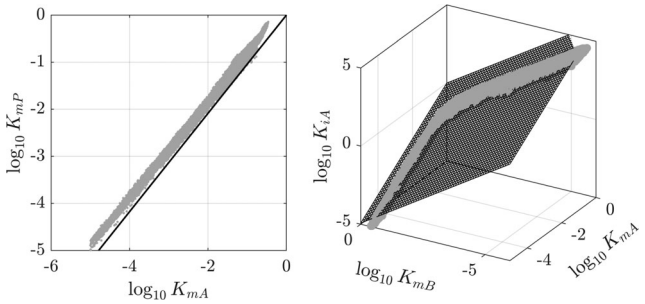
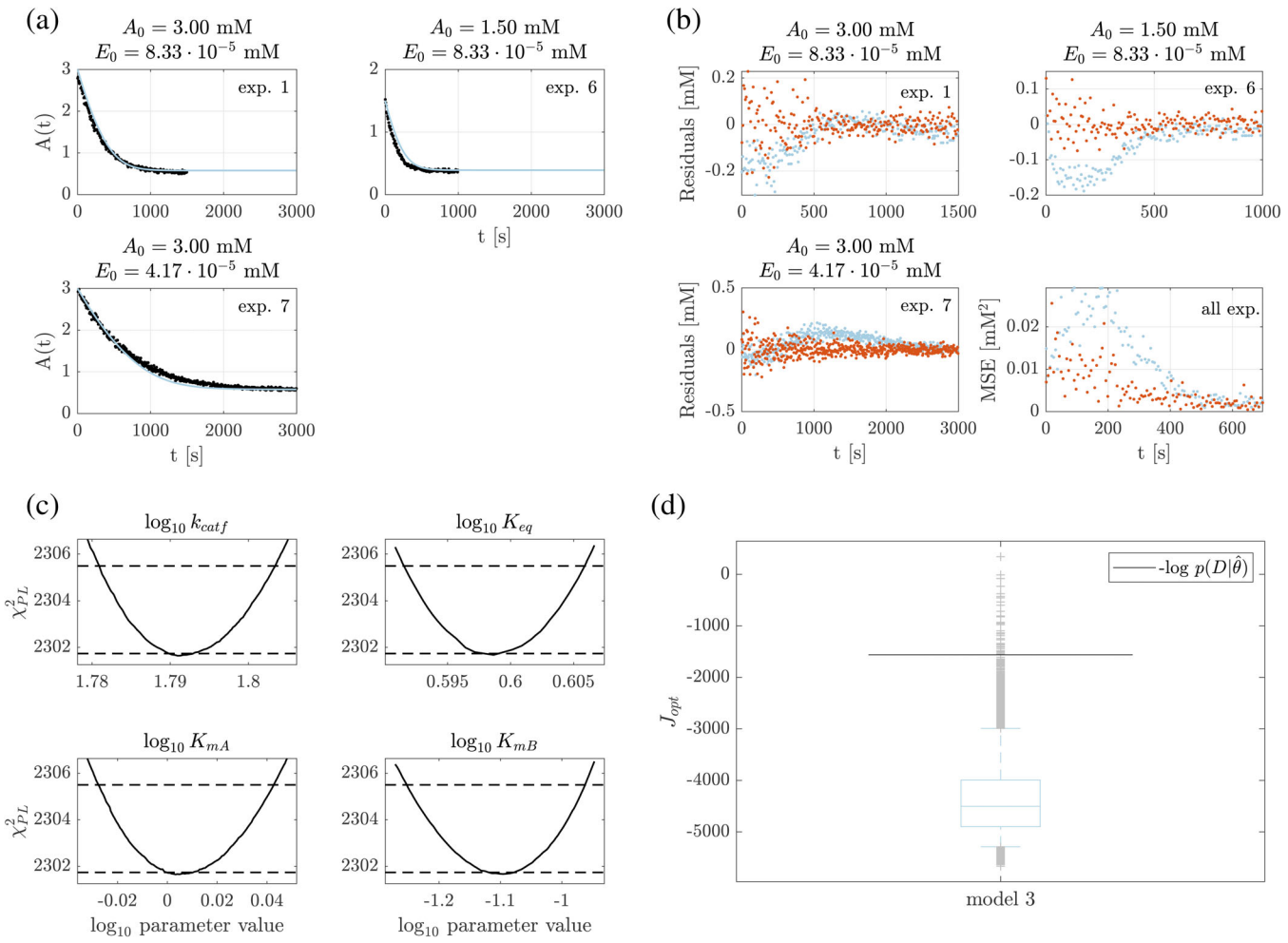


FIGURE 8 Correlation analysis of Model 2. 2D and 3D scatterplots (gray) with their corresponding regression line/plane (black) of the correlated parameters $\log_{10}K_{mP}$ and $\log_{10}K_{mA}$ (left) and $\log_{10}K_{mA}$, $\log_{10}K_{mB}$, and $\log_{10}K_{iA}$ (right), respectively



E

	k_{catf} [s^{-1}]	K_{eq} [mM^{-1}]	K_{mA} [mM]	K_{mB} [mM]	k_{insS} [$mM^{-1}s^{-1}$]
$\log_{10}$-value	0.02	0.01	0.07	0.29	no result
real value	3.21	0.13	0.16	0.05	no result

FIGURE 9 Model evaluation Model 3. (a) Model trajectories, (b) residual, (c) profile likelihood, and (d) bootstrap analysis as well as (e) profile likelihood confidence interval lengths. Analysis was done analogous to Models 1 and 2

between model and data at several places. Residuals simulated by the calibrated model and residuals of the experimental data show significant differences in their distribution, with a worse model fit after reduction. Similar to the MCMC parameter marginals, four of the five parameters have small confidence intervals according to their profile likelihoods (Figure 9c). Calculations of the profile likelihood of K_{insS} did not converge, hence was omitted here. The bootstrap analysis (Figure 9d) confirms results from the residual analysis. The J_{opt} value of the real experiments (black line) lies in the upper tail of the distribution of objective function values estimated by a bootstrap, indicating that the observed deviations are systematic and unlikely to be mere noise.

MAP estimates of the parameters of Model 3 are listed in Figure 10a and are shown as white dots in Figure 10b together with parameter samples of the posterior distributions as 2D scatterplots (lower right half) and corresponding binning plots on a hexagonal lattice (upper left half). 1D parameter marginals are depicted below. Values for k_{catf} and K_{eq} are similar to those of Models 1 and 2, indicating that the data do contain sufficient information to identify these two parameters reliably and that they are not much affected by the differences between the three models. Estimates for the other three parameters differ considerably from respective values of both Models 1 and 2. Confidence intervals resulting from the profile likelihood analysis are given in Figure 9e. Also here, results are in accordance with posterior marginals estimated from the MCMC sampling. Results for additional experiments and profile likelihood together with parameter changes are depicted in Figure S5.

In summary, the reduced model is inferior to Model 2 in terms of Model fit and overall plausibility. This overall conclusion is further supported by statistical measures for model comparison such as the Akaike or the Bayesian information criteria (Table S3).

(a)

Model 3	k_{catf} [s^{-1}]	K_{eq} [mM^{-1}]	K_{mA} [mM]	K_{mB} [mM]	k_{insS} [$mM^{-1}s^{-1}$]
$\log_{10}$-value	1.70	0.56	0.04	−1.08	−2.04
real value	50.27	3.63	1.08	0.08	0.01

(b)

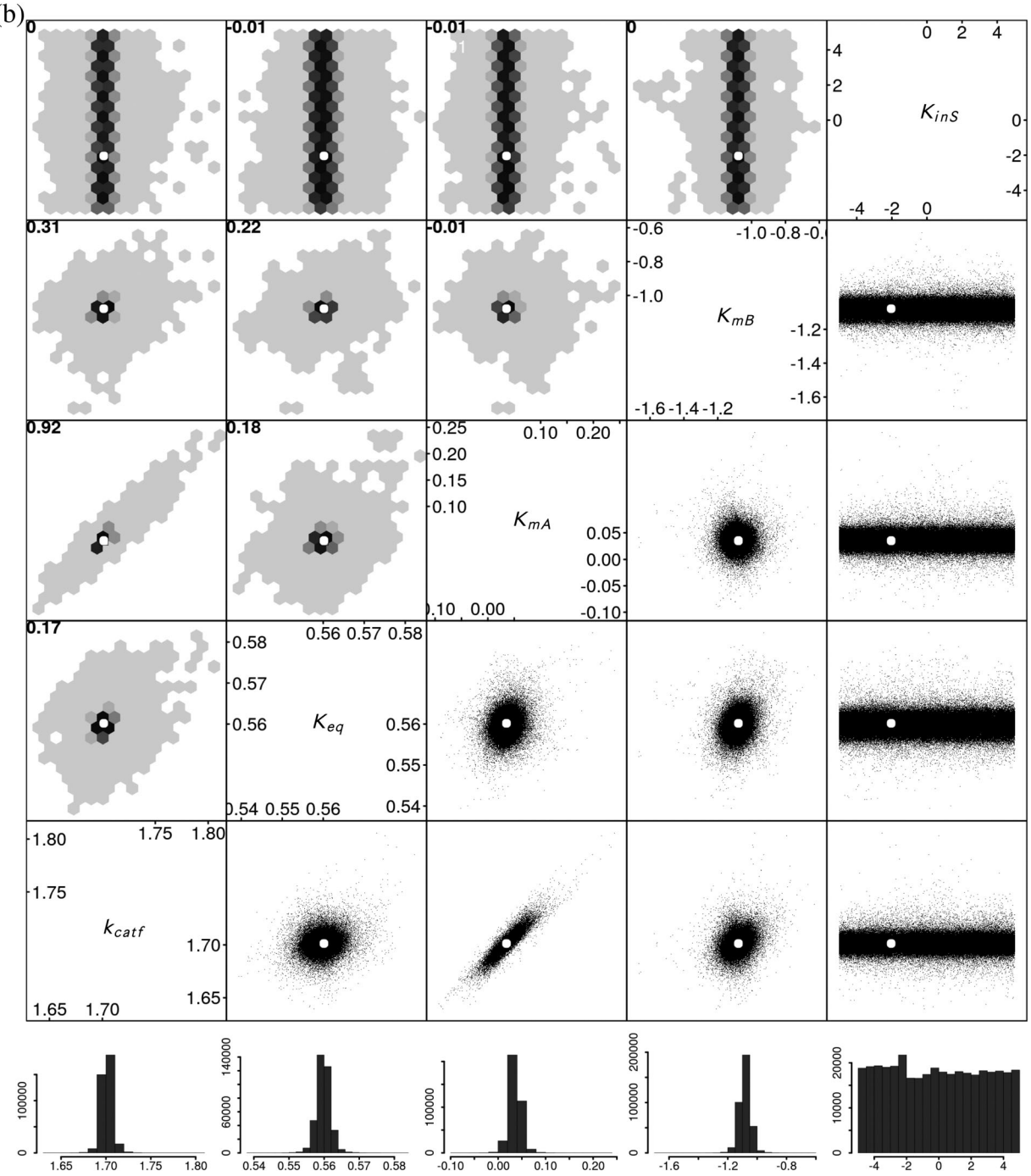


FIGURE 10 Maximum a posteriori (MAP) estimates and 2D marginal posterior densities for Model 3. (a) Best parameter vectors $\hat{\theta}^{MAP}$ of the Markov Chain Monte Carlo (MCMC) sampling for Model 3. (b) Parameter scatterplots (lower right half) and hexbin plots (upper left half) for the MCMC samples of Model 3. MAP estimates $\hat{\theta}^{MAP}$ are indicated by white dots. 1D marginal densities are shown below. Lower and upper bounds have been set to −5 and 5 for log values of all parameters. Units are as listed in (a)

3 | CONCLUSION

In this study, we have introduced a workflow for model calibration, selection and model reduction based on statistical Bayesian sampling, which was exemplified on the symmetric carboligation of 3,5-DMBA to (*R*)-3,3′,5,5′-TMB catalyzed by BAL. Beyond visualizing model fits, also overall plausibility of the stochastic model was investigated via residual analysis and parametric bootstrapping. Together, these methods allowed us to discriminate between different model variants and to identify the overall most plausible model. In this model, the enzyme is assumed to be inactivated by the substrate, which is in accordance with findings in Ohs et al.[4] and Buchholz et al.[5]

Statistical Bayesian approaches are powerful since they deliver a complete stochastic model which can readily be used to simulate data including noise characteristics in silico and to compare these with the experimental data in various ways. This allows to propagate sparseness and empirical variances in experimental data through uncertainties in model parameters to confidence bounds for model predictions. For our showcase example, the different methods that were used here to evaluate model performance all led to consistent results, which is reassuring. For model calibration, the following key conclusions can be drawn from our findings:

First, the estimation of individual model parameter values is difficult and parameter values have to be taken with care. This is not only caused by high correlations between model parameters and non-identifiabilities, but also by their dependence on model formulations. Although all three model variants provide reasonable fits of trajectories, they differ substantially in most of their parameter values. Exceptions are the two parameters k_{catf} and K_{eq}, whose values were robust across all model variants. The overall reaction velocity is directly proportional to k_{catf}, which is known to be a good prerequisite for identifiability of this parameter from time series data. In case of K_{eq}, we anticipate that this parameter is identifiable since it describes the equilibrium of the reaction, for which we have information from experiments with different initial conditions. A sensitivity-based design of experiments could facilitate identifiability in the parameter space, if this is a desired goal.

Second, Model 1 of our study matches the mechanistic model of Zavrel et al.,[11] while methods for model calibration differ. In both studies parameter values of K_{mA}, K_{mB}, K_{mP}, and K_{iA} cannot be estimated precisely and span several orders of magnitude, while the well identifiable parameters k_{catf} and K_{eq} are in agreement with Zavrel et al.[11] Correlation analysis detects correlations between K_{mA} and K_{iA} as well as K_{mB} and K_{iA} in both investigations, whereas further correlations do not coincide. In Buchholz et al.,[5] several kinetic models describing the carboligation of 3,5-DMBA to (*R*)-3,3′,5,5′-TMB are compared. The authors conclude that a substrate-dependent enzyme inactivation exists, which nicely matches our findings. Additionally, while the enzyme inactivation rate k_{inS} derived in Buchholz et al.[5] is in the same range as that of our Model 2 that is selected here, it is two orders of magnitude smaller than that of our Model 3, that was rejected, which is a further support of our results.

Subsequently, we present the methodological key findings: The superiority of statistical methods for model calibration over classical least squares approaches lies in the fact that they include information about the data generation process by taking the noise characteristics in the data into account. This enables a thorough analysis including uncertainties and confidence bounds on all levels, which ultimately also allows to judge overall model plausibility.

Sampling-based approaches as applied here are computationally expensive (for runtimes, see Table S3), which constitutes a clear limitation. Despite many attempts and progress, the development of advanced sampling schemes to improve scalability of these approaches is still a current research topic.[34] Usually, many parameter samples are needed for convergence of the sampling schemes, which is, for example, caused by low acceptance rates of standard sampling schemes especially in cases where parameter correlations are high,[25,26] or because the posterior mass is spread and a large space has to be explored. In addition, if, for example, the data set contains many experiments, evaluation of the likelihood function via numerical integration for each parameter sample might be time consuming as well.

In other studies, we have observed that the parametric bootstrap statistics, as presented here, is highly influenced by the error model, which therefore has to be chosen carefully. Here, we have used a multiplicative error model with predefined SD. Residual analysis indicated that this was a good choice, but we still lack suitable methods for setting up a good error model in general.

Furthermore, we have used linear correlation analysis for model reduction. This was justified by the fact that we observed high correlations between pairs of parameters in the scatterplot analysis. Similar analysis methods that are able to detect nonlinear relations between model parameters exist.[35] In general, however, such purely data-based model reduction techniques are more difficult to interpret and we generally recommend using these techniques that do not deviate much from the underlying physical process if this is possible.

Overall, we are convinced that Bayesian methods for the analysis of dynamic models will become a standard approach once computation times are not that limiting anymore. They are more and more frequently used in different contexts for model calibration, model selection, and uncertainty quantification, see, for example, Davies et al.[36] and Luzyanina and Bocharov.[37]

Future work includes the development of bootstrap methods which exploit information of the full posterior distribution rather than just using the MAP estimate as well as methodology on how to make decisions in case that results of the presented analysis methods are not as consistent as they were in this study. Moreover, in future interdisciplinary projects in which we acquire new data we intend to explicitly integrate optimal experiment design methods into our workflow that go beyond local methods (see, e.g., Stigter et al.[6]) and that make use of the Bayesian viewpoint.

ACKNOWLEDGMENTS

A.J. and N.E.R. acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Initiative—EXC 310/2—and under Germany's Excellence Strategy—EXC 2075-390740016. P.CF.B. and J.P. acknowledge

funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) by grant PL145/16-1.

ORCID

Nicole E. Radde https://orcid.org/0000-0002-5145-0058

REFERENCES

1. Al-Haque N, Santacoloma PA, Neto W, Tufvesson P, Gani R, Woodley JM. A robust methodology for kinetic model parameter estimation for biocatalytic reactions. *Biotechnol Prog*. 2012;28(5):1186-1196.
2. Blackmond DG. Reaction progress kinetic analysis: a powerful methodology for mechanistic studies of complex catalytic reactions. *Angew Chem Int Ed*. 2005;44(28):4302-4320.
3. Marquardt W. Model-based experimental analysis of kinetic phenomena in multi-phase reactive systems. *Chem Eng Res Des*. 2005;83(6):561-573.
4. Ohs R, Leipnitz M, Schöpping M, Spiess AC. Simultaneous identification of reaction and inactivation kinetics of an enzyme-catalyzed carboligation. *Biotechnol Prog*. 2018;34(5):1081-1092.
5. Buchholz PFC, Ohs R, Spiess A, Pleiss J. Progress curve analysis within BioCatNet: comparing kinetic Models for enzyme-catalyzed self-ligation. *Biotechnol J*. 2018;14:1800183.
6. Stigter H, Molenaar J, Joubert D. Observability of complex systems: finding the gap. *Sci Rep*. 2017;7(16566):1-9.
7. Kügler P, Gaubitzer E, Müller S. Parameter identification for chemical reaction systems using sparsity enforcing regularization: a case study for the chlorite-iodide reaction. *J Phys Chem*. 2009;113(12):2775-2785.
8. Kreutz C, Raue A, Kaschek D, Timmer J. Profile likelihood in systems biology. *FEBS J*. 2013;280(11):2564-2571.
9. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian data analysis*. Texts in Statistical Science. 2nd ed. Boca Raton, FL: Chapman & Hall; 2004.
10. Weber P, Hornjik M, Olayioye MA, Hausser A, Radde N. A computational model of PKD and CERT interactions at the trans-Golgi network of mammalian cells. *BMC Syst Biol*. 2015;9:9.
11. Zavrel M, Schmidt T, Michalik C, et al. Mechanistic kinetic model for symmetric carboligations using benzaldehyde lyase. *Biotechnol Bioeng*. 2008;101(1):27-38.
12. Kreutz C, Timmer J. Systems biology: experimental design. *FEBS J*. 2009;276:923-942.
13. Weber P, Kramer A, Dingler C, Radde N. Trajectory-oriented Bayesian experiment design versus fisher A-optimal design: an in depth comparison study. *Bioinformatics*. 2012;28(18):i535-i541.
14. Cleland WW. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: III. Prediction of initial velocity and inhibition patterns by inspection. *Biochim Biophys Acta*. 1963;67:88-196.
15. Dünkelmann P, Kolter-Jung D, Nitsche AS, et al. Development of a donor-acceptor concept for enzymatic cross-coupling reactions of aldehydes: the first asymmetric cross-benzoin condensation. *J Am Chem Soc*. 2002;124(41):12084-12085.
16. Rosche B, Breuer M, Hauer B, Rogers PL. Role of pyruvate in enhancing pyruvate decarboxylase stability towards benzaldehyde. *J Biotechnol*. 2005;115(1):91-99.
17. Sandford V, Breuer M, Hauer B, Rogers P, Rosche B. (R)-phenylacetylcarbinol production in aqueous/organic two-phase systems using partially purified pyruvate decarboxylase from Cabdida utilis. *Biotechnol Bioeng*. 2005;91(2):190-198.
18. Satianegara G, Rogers PL, Rosche B. Comparative studies on enzyme preparations and role of cell components for (R)-phenylacetylcarbinol production in a two-phase biotransformation. *Biotechnol Bioeng*. 2006;94(6):1189-1195.
19. Kreutz C, Bartolome RMM, Maiwald T, et al. An error model for protein quantification. *Bioinformatics*. 2007;23(20):2747-2753.
20. Jensch A, Frey Y, Bitschar K, et al. The tumor suppressor protein DLC1 maintains protein kinase D activity and Golgi secretory function. *J Biol Chem*. 2018;293(37):14407-14416.
21. Benford F. The law of anomalous numbers. *Proc Am Philos Soc*. 1938;78:551-572.
22. Diaconis P. The distribution of leading digits and uniform distribution mod 1. *Ann Probab*. 1977;5:72-81.
23. Kreutz C. New concepts for evaluating the performance of computational methods. *IFAC-PapersOnLine*. 2016;49(26):63-70.
24. Villaverde AF, Froehlich F, Weindl D, Hasenauer J, Banga JR. Benchmark optimization methods for parameter estimation in large kinetic models. *Bioinformatics*. 2018;35(5):830-838.
25. Kramer A, Stathopoulus V, Girolami M, Radde N. MCMC CLIB: an advanced MCMC sampling package for ode models with highly correlated parameters. *Bioinformatics*. 2014;30(20):2991-2992.
26. Kramer A, Calderhead B, Radde N. Hamiltonian Monte Carlo methods for efficient parameter estimation in steady state dynamical systems. *BMC Bioinf*. 2014;15(1):253.
27. Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*. 2014;30(10):1440-1448.
28. Meeker WQ, Escobar LA. Teaching about approximate confidence regions based on maximum likelihood estimation. *Am Stat*. 1995;49(1):48-53.
29. Raue A, Kreutz C, Maiwald T, Klingmüller U, Timmer J. Addressing parameter identifiability by model-based experimentation. *IET Syst Biol*. 2011;5(2):120-130.
30. Efron B. Bayesian inference and the parametric bootstrap. *Ann Appl Stat*. 2012;6(4):1971-1997.
31. Fall CP, Marland ES, Wagner JM, Tyson JJ, eds. *Computational cell biology*. Interdisciplinary applied mathematics. Vol 20. New York, NY: Springer; 2005.
32. Murray JD. *Mathematical biology—an introduction*. Interdisciplinary applied mathematics. Vol 17. Berlin, Germany: Springer; 2002.
33. Segel LA. *Modeling dynamic phenomena in molecular and cellular biology*. Cambridge, UK: Cambridge University Press; 1984.
34. Robert CP, Elvira V, Tawn N, Wu C. Accelerating MCMC algorithms. *WIREs Comput Stat*. 2018;10:e1435.
35. Hengl S, Kreutz C, Timmer J, Maiwald T. Data-based identifiability analysis of non -linear dynamical models. *Bioinformatics*. 2007;23(19):2612-2618.
36. Davies V, Harvey WT, Reeve R, Husmeier D. Improving the identification of antigenic sites in the H1N1 influenza virus through accounting for the experimental structure in a sparse hierarchical Bayesian model. *J Roy Stat Soc C*. 2019;68(4):859-885.
37. Luzyanina T, Bocharov G. Markov chain Monte Carlo parameter estimation of the ODE compartmental cell growth model. *Math Biol Bioinf*. 2018;13(2):376-391.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Eisenkolb I, Jensch A, Eisenkolb K, et al. Modeling of biocatalytic reactions: A workflow for model calibration, selection, and validation using Bayesian statistics. *AIChE J*. 2019;e16866. https://doi.org/10.1002/aic.16866